Optimal Human Capital Policies*

Radim Bohacek, Economics Institute and Charles University

Marek Kapicka, University of California, Santa Barbara†

May 13, 2007

Abstract

We analyze optimal income taxes and optimal schooling subsidies in a dynamic private information economy with observable human capital accumulation. We show that the marginal schooling subsidies are, under plausible conditions, positive and that they are zero at both endpoints of the skill distribution. For our utility specification, we suggest that the optimum can be implemented in a competitive equilibrium with a tax system where taxes each period depend only on current income and current schooling.

We calibrate the model to the U.S. economy in order to quantify the optimal policies and evaluate their impact in the transition and in the steady state. We find that the optimal schooling policies are significantly smaller than the optimal marginal income taxes. If their are introduced jointly with optimal income taxes their welfare and

*We would like to thank the participants of the Carnegie-Rochester conference in April 2007 and especially Borys Grochulski for helpful comments. This research has originated during out stay at the Univeristy of Tokyo. We would like to thank Toni Braun for his hospitality during our stay.

†Corresponding Author. Email: mkapicka@econ.ucsb.edu. Address: 2127 North Hall, Univesity of California, Santa Barbara, CA 93106.

aggregate effects are small. However, if the income taxes are not set optimally then the optimal schooling policies have significant welfare and aggregate effects.

1 Introduction

We analyze optimal government policies in an economy where agents are heterogeneous in their abilities to work, but can also invest in human capital to augment their skills. We assume that their abilities are their private information. The fact that abilities are unobservable by the government creates a friction that provides a foundation for optimal income taxes and optimal educational subsidies.

If there are no informational frictions, it is optimal for the allocations to satisfy two conditions. The first one equates the marginal rate of substitution between consumption and leisure to the wage rate. The second one equates marginal costs of investment in human capital to the marginal benefits of doing so. The second condition is expressed in terms of an Euler equation for human capital. Static models with private information show that the efficient allocation creates a wedge between the wage and the marginal rate of substitution. In this paper we analyze whether private information also creates distortions in the Euler equation for human capital. To do so, we define an intertemporal human capital wedge in the Euler equation as a wedge between the marginal costs of investment in human capital and the marginal benefits of such investment. We find that it is efficient to have positive human capital wedge as long as it is optimal to increase the agent's non-leisure time and labor supply relative to schooling over the transition. A positive human capital wedge means that the marginal cost of investment in human capital is, in the optimum, larger than the marginal benefit. We interpret the optimal human capital wedge as a foundation for positive schooling policy subsidies.

The intuition behind the positive schooling subsidies is that since marginal income taxes are positive, labor supply is distorted downwards, compared to the first best labor supply. In the absence of schooling subsidies, schooling is undistorted relative to labor supply and hence it is distorted downwards compared to the first best schooling. By introducing positive schooling subsidies, the government reduces the schooling distortion at the expense of introducing a new distortion of schooling relative to labor supply. We show that such trade-off is in general welfare improving.

Similarly to Kapicka [19], we consider a decentralization in a competitive equilibrium where the efficient allocations are implemented by allowing the agents to borrow and lend in a credit market and by using a tax system where in each period taxes depend only on current income and current schooling, although they are not additively separable. We do not have a proof that this tax system is always sufficient to implement the optimum. Instead, we present a numerical implementation verification procedure that checks if the efficient allocation can be implemented by such a tax system. We find that for our efficient allocation the implementation verification procedure is successful.

We provide quantitative characterization of the optimal policies. Numerical simulations show that the optimal marginal income tax schedule varies very little over time. For majority of the population the optimal marginal income taxes lie between 40-65%. The marginal schooling subsidy schedule also changes very little with time. Unlike the marginal tax rates, the optimal marginal schooling subsidy is relatively small, lying between 0-7% for most of the population.

To isolate the effects of schooling subsidies from the effects of optimal income taxes, we also consider two alternative, more restricted policy reforms. We first exogenously restrict the schooling policies to be zero. We find that the quantitative effects of optimal schooling subsidies are very small: the welfare gain from introducing schooling subsidies on top of the optimal income tax schedule is only 0.3% in terms of consumption equivalents. Similarly, the effect on aggregate output is small.

We then study optimal schooling policies when the income tax policy is exogenously given by a flat tax of 40%. We find that in such case the welfare gain from schooling subsidies is significantly larger, being 8.8% in terms of consumption equivalents. The aggregate effects of optimal schooling policies are also more significant: When schooling subsidies are used optimally given the flat income tax, the aggregate steady state output is about 9.9% higher than if only the flat income tax is used. These results imply that schooling subsidies may be an important policy tool and can significantly increase welfare if the income tax policies are not optimal.

1.1 Relation to the Literature

This paper connects two strands of literature. The first one builds on Mirrlees's [23] framework where private information about one's abilities provides a foundation for optimal income taxes. This framework has been recently extended to dynamic environments with physical capital by Golosov, Kocherlakota, Tsyvinski [10], Kocherlakota [21], Werning [31] and Albanesi and Sleet [2]. One of their main findings, that capital income taxes should be zero only if individual's abilities are permanent is reflected in our implementation where the social planner allows the agents to borrow and save without any restrictions. In their case, optimal allocations exhibit a positive intertemporal wedge between the marginal costs of saving and marginal benefits of doing so. Although this wedge is quite distinct from our intertemporal human capital wedge, the intuition why both wedges are positive is similar. A positive intertemporal wedge (and positive capital taxes) exists in order to decrease incentives to save. Lower savings is beneficial because it induces higher labor supply in the future. A positive human capital intertemporal wedge (and positive schooling subsidies) exists in order to increase human capital investment. Higher human capital investment increases the marginal product of labor in the future and therefore induces higher labor supply in the future, similarly to lower savings.

Dynamic private information economies with endogenous human capital formation were considered by Kapicka [18] and [19] where the setting was simplified by the fact that human capital was assumed unobservable or, equivalently, no schooling policies were feasible. This paper extends these results by assuming that human capital is observable and thus a richer set of government policies is allowed. Grochulski and Piskorski [12] study a two-period economy where agent's skills are also affected by human capital accumulation. Their environment allows for more general ability shocks but since they also assume that human capital is unobservable, no direct schooling policies are again feasible. Our paper is also close to Bovenberg and Jacobs [4] who look at optimal educational subsidies in a static framework. Unlike them, we assume that the major cost of obtaining education is one's time, not physical resources.

This paper also connects to the second strand of literature, which analyzes the determinants of human capital accumulation in a general equilibrium framework. These models naturally have much richer environment than our model and we build on them in several ways. Following Huggett, Ventura and Yaron [16] and Andolfatto, Ferrall and Gomme [3], we focus on time and not goods as the main cost of increasing human capital. Huggett, Ventura and Yaron also stress that permanent ability differences are essential to reproduce the observed facts about human capital and earning dynamics and our model is consistent with this fact.

Several papers have studied the effects of partial tax reforms on the accumulation of human capital and labor supply. Although these models are again richer in their setup, the set of tax instruments considered is usually restricted to linear taxes on current variables and no explicit motivation for these restrictions is provided. Heckman, Lochner and Taber [15] and Taber [29] analyze the effects of a flat tax reform in both partial and general equilibrium setup. Similarly to them, we find large partial equilibrium effects of all income tax reforms on human capital accumulation. One of the main conclusions reached by this literature, discussed also in Carneiro and Heckman [6], is that the current tax system does not provide too many incentives for human capital investment. This conclusion is reinforced by our findings that even when the income taxes are flat there is a significant space for additional schooling policies to encourage human capital accumulation. Only when income taxes are set optimally is the space for human capital policies small.

The paper is organized as follows. The next section presents the model. Section 3 characterizes the efficient allocations. Section 4 defines the recursive problem for efficient time allocations. Implementation in a market economy is in section 5. Section 6 shows the results from numerical simulation and section 7 concludes. All proofs and computational details are in the Appendix.

2 Setup

Time is discrete, $t \geq 0$. There is a measure 1 of agents in the economy. Each individual is associated with an ability level $\theta \in [0, \bar{\theta}] = \Theta$. We assume that the ability level does not change over time. Although this assumption is clearly restrictive, it keeps the model tractable and is consistent with recent empirical evidence that income processes are very persistent[1]. Distribution of abilities is given by a distribution F. We assume that F is twice differentiable and has density $f(\theta)$. The abilities are private information of each agent. The abilities affect earnings of the agent in a way specified below.

Each agent is endowed with one unit of time. In each period, time can be divided between leisure, work, and time spent by human capital accumulation. Denote working time as l_t and time spent by accumulating human capital by s_t.[2]

Period utility of each person depends on consumption $c_t \geq 0$ and nonleisure time $l_t + s_t \in [0, 1]$. We assume that the utility function is such that the income effects on leisure are zero: an individual evaluates consumption and leisure sequences according to

$$\sum_{t=0}^{\infty} \beta^t [c_t - v(l_t + s_t)] \qquad 0 < \beta < 1,$$

where $v : [0, 1] \to \mathbf{R}$ is utility from leisure. We assume that v is continuously differentiable on $\mathbf{R}_{++} \times (0, 1]$, strictly increasing, strictly convex and bounded by above by some constant $\bar{v}$. We also define a period utility $u_t = c_t - v(l_t + s_t)$ since in our analysis it will become more

[1] Heathcote, Storesletten and Violante [13] find that the autocorrelation of wage shocks is about 0.94, and thus exhibits near random walk behavior.

[2] It is necessary to interpret time spent by accumulating human capital quite extensively. This does not include only time spent in schools but also other activities that increase individual's human capital, i.e. on-the-job training.

convenient to work with period utility rather than consumption.

Each individual starts with initial human capital h_0. To reduce the complexity of the model we assume that h_0 is identical for all people and is observed by the government. Agent's human capital at the beginning of period $t+1$ is denoted h_{t+1}. It depends on time spent accumulating it in previous period s_t, previous level of human capital h_t and is given by a human capital accumulation function $G : \mathbf{R}_+ \times [0,1] \rightarrow \mathbf{R}_+$:

$$h_{t+1} = G(h_t, s_t).$$

Several assumptions are made about G. It is assumed to be continuously differentiable on $\mathbf{R}_{++} \times (0,1]$, strictly increasing and strictly concave. There is an upper bound on human capital stock $\bar{h}$ that cannot be exceeded even if all time is invested in schooling: $G(\bar{h}, 1) \leq \bar{h}$. If, on the other hand, no additional human capital investments are made then human capital depreciates over time: $G(h,0) < h$ if $h \in (0, \bar{h}]$. It is also assumed that $\lim_{s \to 0} G_s(s,h) = +\infty$ for all $h \in (0, \bar{h}]$ and that G_h and $\frac{1}{G_s}$ are bounded on $\left[0, \bar{h}\right] \times [0,1]$.

Human capital is accumulated in order to increase the skills of the agent. We assume an efficiency unit specification: a person with human capital h_t and abilities θ has skills θh_t. If she works l_t hours, she produces output $y_t = \theta h_t l_t$.

It is assumed that the government can borrow or lend at an interest rate $\frac{1}{\beta} - 1$. The government has to finance an exogenous sequence of expenditures with present value E. The government is supposed to maximize the social welfare function

$$\int_{\Theta} \sum_{t=0}^{\infty} \beta^t U(u_t) f d\theta$$

where $U : \mathbf{R}_+ \to \mathbf{R}$ is continuously differentiable on $\mathbf{R}_{++} \times (0, 1]$, strictly increasing and strictly concave. The function U measures the government's aversion toward inequality.

We define an *allocation* to be a sequence of functions $\psi = \{u_t, y_t, h_{t+1}\}_{t=0}^{\infty}$ where $u_t : \Theta \to \mathbf{R}_+$ specifies consumption in period t, $y_t : \Theta \to \mathbf{R}_+$ specifies output in period t and $h_{t+1} : \Theta \to \mathbf{R}_+$ specifies human capital at the beginning of period $t+1$.[3] The allocation must satisfy two feasibility constraints. The first one is resource feasibility. Present discounted value of aggregate consumption and government expenditures must be smaller or equal to the present discounted value of aggregate output. Using period utility rather than consumption, the resource constraint becomes

$$E + \int_{\Theta} \sum_{t=0}^{\infty} \beta^t u_t(\theta) f(\theta) d\theta \leq \int_{\Theta} \sum_{t=0}^{\infty} \beta^t [y_t(\theta) - v(\frac{y_t(\theta)}{\theta h_t(\theta)} + S(h_t(\theta), h_{t+1}(\theta))] f(\theta) d\theta. \qquad (1)$$

where $S(h_t, h_{t+1})$ is schooling time required to have next period human capital h_{t+1} when current human capital is h_t. The second constraint is feasibility of time allocations for each agent in each period:

$$0 \leq \frac{y_t(\theta)}{\theta h_t(\theta)} + S(h_t(\theta), h_{t+1}(\theta)) \leq 1 \qquad \forall \theta \in \Theta, \forall t \geq 0,$$

[3]Schooling s_t and labor supply l_t can be recovered by inverting the human capital production function and output production function respectively.

Since the agent's ability level is a private information, the social planner needs to elicit agent's type from her. At the beginning of period 0 the agent is asked to report her type to the social planner. The utility of a θ - type agent who reports $\hat{\theta}$ is given by

$$V_\psi(h_0, \theta; \hat{\theta}) = \sum_{t \geq 0} \beta^t [u_t(\hat{\theta}) + v(\frac{y_t(\hat{\theta})}{\hat{\theta} h_t(\hat{\theta})} + S(h_t(\hat{\theta}), h_{t+1}(\hat{\theta})) - v(\frac{y_t(\hat{\theta})}{\theta h_t(\hat{\theta})} + S(h_t(\hat{\theta}), h_{t+1}(\hat{\theta}))].$$

The incentive compatibility constraint requires the allocation to be such that θ - type agent prefers to report her own type to any other report: for all $\theta \in \Theta$,

$$V_\psi(h_0, \theta; \theta) \geq V_\psi(h_0, \theta; \hat{\theta}) \quad \forall \hat{\theta} \in \Theta. \tag{2}$$

An allocation that is both feasible and incentive compatible will be called *incentive-feasible*. Denote the set of all incentive feasible allocations by Ψ^{IF}. The social planner chooses an incentive feasible allocation to maximize the social welfare function:

$$W(h_0, E) = \max_{\psi \in \Psi^{IF}} \int_\Theta U(u_t(\theta)) f(\theta) d\theta, \tag{3}$$

where $W(h_0, E)$ is the value of the social planner's problem when the initial human capital is h_0 and the present value of government expenditures is E. The *efficient allocation* is the allocation that attains the maximum of (3).

2.1 First Order Approach

We will now simplify the incentive compatibility constraint by using a first order approach. In what follows, it will become convenient to redefine allocation by replacing income $y_t(\theta)$ by labor supply $l_t(\theta) = \frac{y_t(\theta)}{\theta h_t(\theta)}$. An allocation $\psi = \{u_t, l_t, h_{t+1}\}_{t=0}^{\infty}$ now consists of sequences of period utility, labor supply and human capital. A necessary condition for an allocation to be incentive compatible is given by the envelope condition:

$$\sum_{t=0}^{\infty} \beta^t u_t(\theta) = \int_{\underline{\theta}}^{\theta} \sum_{t=0}^{\infty} \beta^t v_t'(\varepsilon) l_t(\varepsilon) \frac{d\varepsilon}{\varepsilon} + u_0, \tag{4}$$

where $v_t'(\varepsilon) = v'[l_t(\varepsilon) + S(h_t(\varepsilon), h_{t+1}(\varepsilon)]$. Equation (4) shows how the agent's period utility varies with her type. The variation in period utility is proportional to the informational rent the agent obtains from having a given ability level. The utility of the lowest type agent, who gets no informational rent, is given by u_0.

The envelope condition (4) is necessary for an allocation to be incentive compatible. Next Lemma shows that the envelope condition (4) is also sufficient, provided that both schooling $s_t(\theta)$ and income to human capital ratio $\frac{y_t(\theta)}{h_t(\theta)}$ are increasing in θ in all periods. The proof is relegated to the Appendix.

Lemma 1 *Suppose that $s_t(\theta)$ and $\frac{y_t(\theta)}{h_t(\theta)}$ are increasing in θ for all $t \geq 0$. Then an allocation is incentive-compatible if and only if it satisfies (4).*

The monotonicity conditions of Lemma (**??**), together with the envelope condition (4) are enough to establish the validity of the first order approach. Note however that the monotonicity conditions of this Lemma are not necessary. If they fail in some periods, an allocation satisfying (4) might still be incentive compatible, but one needs to find other ways to verify it.[4].

3 Efficient Allocations

Let λ be the Lagrange multiplier on the resource constraint (1) and $\mu(\theta)f(\theta)$ be the Lagrange multiplier on the envelope condition (4). The Lagrangean for the social planner's problem (3) is

$$\mathcal{L}(\psi, \lambda, \mu) = \int_{\Theta} \sum_{t=0}^{\infty} \beta^t \{U(u_t) + \lambda[\theta h_t l_t - v_t - u] - \mu[\int_{\underline{\theta}}^{\theta} v_t' l_t \frac{d\varepsilon}{\varepsilon} + u_0 - u_t]\} f d\theta - \lambda E.$$

Let $\psi_{l,h} = \{l_t, h_{t+1}\}_{t=0}^{\infty}$ be a section of an allocation consisting only of labor supply and human capital sequences. Define Π to be the set of all labor supply and human capital sequences $\psi_{l,h}$ such that the implied leisure is feasible in all periods: $\psi_{l,h} \in \Pi$ if for all $\theta \in \Theta$ and all $t \geq 0$, $0 \leq S(h_t(\theta), h_{t+1}(\theta)) \leq 1 - l_t(\theta)$.[5] The efficient allocation is then the saddle point of the Lagrangean:

$$W(E, h_0) = \max_{\psi} \min_{\lambda,\mu} \mathcal{L}(\psi, \lambda, \mu) \qquad s.t.\ \psi_{l,h} \in \Pi. \tag{5}$$

[4]One way to do it is to devise an ex-post incentive compatibility verification procedure similar to the one suggested by Abraham and Pavoni [1].

[5]Writing $\psi_{l,h} \in \Pi$ means that this property holds for all $\theta \in \Theta$.

We now partially characterize the optimum. The first-order conditions in $u_t(\theta)$ and in u_0 are

$$U'(u_t) = \lambda - \mu, \tag{6}$$

$$\int_{\Theta} \mu f d\theta = 0. \tag{7}$$

The first order condition (6) implies immediately that in the optimum, the period utility is constant over time. To simplify notation, we will therefore drop the time subscript on period utility. Combining both first-order conditions together, one obtains that the efficient allocations must satisfy

$$\int_{\Theta} [U'(u) - \lambda] f d\theta = 0. \tag{8}$$

This equation says that, on average, the marginal utility must be equal to the shadow price of resources. The reason is that increasing or decreasing utility uniformly for all agents is feasible for the social planner and has no effect on the incentive compatibility constraint. But private information introduces distortions in a sense that marginal utility is not equal to the shadow price of resources agent by agent: The Lagrange multiplier μ is nonzero for almost all agents.

We now define a partially optimized Lagrangean $\hat{\mathcal{L}}_{u,\lambda}(\psi_{l,h})$ where (6) and (7) are assumed to hold.[6] The Lagrange multiplier on the envelope condition μ is eliminated directly by using

[6]Since (6) and (7) hold only in the optimum, the value of $\hat{\mathcal{L}}_{u,\lambda}(\psi_{l,h})$ is not necessarily always identical to the value of the Lagrangean $\mathcal{L}$ evaluated at the same values. What matters, however, is that the two values are the same at the optimum.

(6) and (7). The partially optimized Lagrangean is

$$\hat{\mathcal{L}}_{u,\lambda}(\psi_{l,h}) = \int_{\Theta} \sum_{t=0}^{\infty} \beta^t \{U(u) + \lambda[\theta h_t l_t - v_t - u] + [U'(u) - \lambda][\int_{\underline{\theta}}^{\theta} v_t' l_t \frac{d\varepsilon}{\varepsilon} - u]\} f d\theta - \lambda E$$

$$= \int_{\Theta} \sum_{t=0}^{\infty} \beta^t \{\lambda[\theta h_t l_t - v_t] - v_t' l_t \frac{1}{f\theta} \int_{\underline{\theta}}^{\theta} [U'(u) - \lambda] f(\varepsilon) d\varepsilon\} f d\theta + \xi(u) - \lambda E \qquad (9)$$

where $\xi(u) = \frac{1}{1-\beta} \int_{\underline{\theta}}^{\overline{\theta}} \{U(u) - U'(u)u\} f d\theta$. The second equality is obtained by integrating by parts and using (7).

Define a *cumulative distortion function* $X_{u,\lambda}(\theta)$ to be a function giving, for each agent, the average percentage deviation of marginal utility from the shadow price, the average being taken across all agents with lower abilities. The cumulative distortion function is

$$X_{u,\lambda}(\theta) = \frac{1}{\lambda} \int_{\underline{\theta}}^{\theta} [U'(u) - \lambda] f d\varepsilon.$$

The cumulative distortion function appears directly in the Lagrangean (9) and will play an important role later on. It is therefore worthwhile to analyze its properties. We show next that it is a hump-shaped nonnegative function that starts and ends at zero.

Lemma 2 *$X_{u,\lambda}(\theta)$ is positive for all $\theta \in \Theta$. In addition, $X_{u,\lambda}(\underline{\theta}) = X_{u,\lambda}(\overline{\theta}) = 0$.*

We will now conjecture and later prove that if u and λ are the optimal values that satisfy

(5) then the social planner's problem can be written as

$$\begin{aligned}
W(E,h_0) &= \max_{\psi_{l,h}} \hat{\mathcal{L}}_{u,\lambda}(\psi_{l,h}) \quad s.t.\ \psi_{l,h} \in \Pi \\
&= \max_{\psi_{l,h}} \lambda \int_{\Theta} \sum_{t=0}^{\infty} \beta^t [\theta h_t l_t - v_t - v_t' l_t \frac{1}{f\theta} X_{u,\lambda}(\theta)] f d\theta + \xi(u) - \lambda E \quad s.t.\ \psi_{l,h} \in \Pi \\
&= \lambda \int_{\Theta} \left\{ \max_{\psi_{l,h}(\theta)\in\Pi} \sum_{t=0}^{\infty} \beta^t [\theta h_t l_t - v_t - v_t' l_t \frac{1}{f\theta} X_{u,\lambda}(\theta)] \right\} f d\theta + \xi(u) - \lambda E.
\end{aligned}$$

Thus, the maximization of the partially optimized Lagrangean can be broken into a continuum of separate maximization problems. These maximization problems are connected only by the cumulative distortion function. In other words, knowledge of $x = X_{u,\lambda}(\theta)$ is sufficient to compute the optimal sequence of labor supply and human capital stock for a $\theta-$type agent. For an arbitrary distortion x, the value of this problem is given by a function $Q_0(x,\theta,h_0)$ satisfying

$$Q_0(x,\theta,h_0) = \max_{\psi_{l,h}\in\Pi} \sum_{t=0}^{\infty} \beta^t \{\theta h_t l_t - v_t - v_t' l_t \frac{x}{f\theta}\}. \tag{10}$$

Denote the solution to this problem by a sequence of functions $\psi_{l,h}^d = \{l_t^d, h_{t+1}^d\}_{t\geq 0}$. One way to look at this problem is to view it as an individual's problem, which is affected by a cumulative distortion.[7] We will therefore call this problem an *individual's distorted problem*. It will be shown that the individual's distorted problem can be conveniently written recursively. Thus, once the cumulative distortion function is known, one can solve fairly easily for the optimal human capital and labor supply sequences - and that is the main benefit of

[7] If $x = 0$ then this problem is identical to the individual's problem with no government intervention.

this formulation.[8]

We will now verify that the results found so far do indeed provide a solution to the relaxed social planner's problem. The proof of the Theorem is relegated to the Appendix.

Theorem 3 (The Decomposition Theorem) *An allocation ψ^*, together with the Lagrange multiplier on the resource constraint λ^*, solves the relaxed social planner's problem if and only if period utility u^* is constant over time for all agents and labor supply and human capital sequences satisfy*

$$\psi^*_{l,h}(\theta, h_0) = \psi^d_{l,h}(X^*(\theta), \theta, h_0)$$

for all $t > 0$ and for all $\theta \in \Theta$, where $X^*(\theta) = \frac{1}{\lambda^*} \int_{\underline{\theta}}^{\theta} [U'(u^*) - \lambda^*] f d\varepsilon$ and satisfies $X^*(\bar{\theta}) = 0$.

Although the social planner's problem is not a convex problem, sufficiency is obtained even with our reliance on the first-order condition in u. The reason is that all the nonconvex elements of the problem appear in the individual's distorted problem for which the first-order conditions were not used.

4 A Bellman Equation for Efficient Time Allocations

The individual's distorted problem involves one endogenous state variable - human capital and two exogenous state variables - individual's ability level θ and the distortion x. Both of the exogenous state variables are constant over time and thus work like a fixed effect. Define

[8] In Section 6 we will describe an iterative procedure that is used to ensure that the optimal cumulative distortion function is found.

a value function $Q_{x,\theta} : [0, \bar{h}] \to \mathbf{R}$ to be a solution to the following Bellman equation:

$$Q_{x,\theta}(h) = \max_{l,h'}\{\theta hl - v - v'l\frac{x}{f\theta} + \beta Q_{x,\theta}(h')\}. \tag{11}$$

We will now show the equivalence between this restricted sequence problem and the dynamic program (11). The period return function $\theta hl + v - v'l\frac{x}{f\theta}$ is bounded above by $\theta\bar{h} + \bar{v}$ since v is bounded above by $\bar{v}$ by assumption and both v' and x are positive. Also, for any $h \in [0, \bar{h}]$ one can choose a pair l, h' in such a way that it satisfies all the constraints of the dynamic program.[9] The assumptions of Theorems 4.2 and 4.4 of Stokey, Lucas and Prescott [27] are therefore satisfied, $Q_0(x, \theta, .)$ satisfies the Bellman equation (11) and the solution to the sequence problem attains its maximum. The converse is also true, as follows from Exercise 4.3 and Theorem 4.5 of [27], and the optimal policy functions that attain maximum of (11) can be used to generate the solution of the sequence problem (10).

4.1 Characterizing Efficient Time Allocations

It is easy to show that in a world with no distortions (i.e. with $x = 0$) or when human capital is unobservable[10], the following Euler equation for schooling must hold:

$$\frac{v'_t}{G_{s_t}} = \beta v'_{t+1}(\frac{l_{t+1}}{h_{t+1}} + \frac{G_{h_{t+1}}}{G_{s_{t+1}}}). \tag{12}$$

Equation (12) equates marginal costs of increasing human capital by one unit with marginal benefits of doing so. Marginal costs are given by the left-hand side of (12). There

[9]Setting h' such that $g(h, h') = 0$, $l \in (0, 1)$ is an example.
[10]See Kapicka [19].

are two types of marginal benefits. First, the productivity increases next period and so less time is needed to produce a given income. This effect is captured by the term $\beta v'_{t+1}\frac{l_{t+1}}{h_{t+1}}$. Second, less schooling is needed to produce future human capital. This benefit is captured by the term $\beta v'_{t+1}\frac{G_{h_{t+1}}}{G_{s_{t+1}}}$.

Since the cumulative distortion function is typically nonzero, the equation (12) will in general be violated. That is, the social planner will want to create a wedge between marginal costs of investment in human capital and marginal benefits. We will call this wedge an *intertemporal human capital wedge*. Formally, the intertemporal human capital wedge $\Delta_t : \Theta \to R_+$ is defined as as follows:

$$\Delta_t(\theta) \equiv \frac{v'_t}{G_{s_t}} - \beta v'_{t+1}(\frac{l_{t+1}}{h_{t+1}} + \frac{G_{h_{t+1}}}{G_{s_{t+1}}}). \tag{13}$$

If, in the optimum, marginal costs of investment in human capital exceed marginal benefits, the intertemporal wedge is positive. If the opposit is true, the intertemporal wedge is negative.and is negative.

We will now discuss if the efficient human capital wedge is positive or negative. First-order conditions for the Bellman equation (11) imply[11] that labor supply and human capital be related over time in the following way:

$$(1+\omega_{t,t+1})\frac{v'_t}{G_{s_t}} = \beta v'_{t+1}(\frac{l_{t+1}}{h_{t+1} - v'_{t+1}\frac{x}{f\theta^2}} + \frac{G_{h_{t+1}}}{G_{s_{t+1}}}), \tag{14}$$

[11] We assume that the labor supply is strictly positive.

where

$$\omega_{t,t+1} = \frac{1 + \frac{x}{f\theta}\frac{l_t}{l_t+s_t}\rho(l_t + s_t)}{1 + \frac{x}{f\theta}\frac{l_{t+1}}{l_{t+1}+s_{t+1}}\rho(l_{t+1} + s_{t+1})} - 1,$$

where $\rho(n) = \frac{v''(n)}{v'(n)}n$. The equation (14) shows that the efficient time allocations in general depart from the Euler Equation (12) in two ways. First the left-hand side contains the term $\omega_{t,t+1}$. Second, l_{t+1} is on the right-hand side is divided by $h_{t+1} - v'_{t+1}\frac{x}{f\theta^2}$ rather than h_{t+1} itself.

Suppose for now that $\omega_{t,t+1} = 0$. Since $v'_{t+1}\frac{x}{f\theta^2}$ is surely positive, the efficient human capital wedge must be positive as well. It is more complicated to determine the sign of $\omega_{t,t+1}$ since the value function is not necessarily concave and so schooling and labor supply and next period human capital may not be monotone in current human capital. Negative value of $\omega_{t,t+1}$ will ensure that the human capital wedge is positive. If, on the other hand, $\omega_{t,t+1}$ is greater then zero then one cannot determine the sign of the wedge.

We have the following result that states sufficient conditions for the human capital wedge to be positive. It is easily proven by showing that, under its assumptions, $\omega_{t,t+1}$ is negative.

Proposition 4 *Suppose that $\rho(n)$ is increasing in n. If both $l_t + s_t$ and $\frac{l_t}{l_t+s_t}$ increase over time then the human capital wedge is positive.*

The fraction of nonleisure time devoted to labor is more likely to be increasing over time if an agent increases her human capital over time. The reason is that more human capital implies higher rates of return from working, while less schooling is needed to achieve a given level of next period human capital. Nonleisure time will increase if the positive effect on labor supply outweights the negative effect on schooling. In such case, agents who accumulate human capital over the transition path will experience a positive human capital wedge.

It is easy to see that in a steady state $\omega_{t,t+1} = 0$. We therefore have the following corollary:

Corollary 5 *The human capital wedge is always positive in a steady state.*

Thus, although one may obtain a different result along the transition path, especially for agents who start with human capital that is too high, the human capital wedge will turn positive in a steady state. Since the human capital wedge is also positive when $\omega_{t,t+1} > 0$ but small, the same result will be found near the steady state.

One should therefore expect that, quantitatively, the human capital wedge will be positive. If this is the case then marginal costs of investment in human capital exceed marginal benefits. Thus, the social planner forces the agents to accumulate more human capital, relative to the labor supply, than the agents would accumulate if the human capital were unobservable. The intuition for this result is, that compared to the first best allocation the social planner typically distorts individual's labor supply downwards. If human capital is unobservable then schooling is also distorted downwards, because rates of return from investment in human capital decrease. If, however, human capital is observable, then the social planner can reduce the second inefficiency by forcing the agents to increase their schooling relatively to their labor supply.

Since by Lemma 2 the cumulative distortion is zero at both endpoints of the ability distribution, it follows from (14) that the same must be true about the intertemporal human capital wedge at both endpoints:

Lemma 6 *The intertemporal human capital wedge is zero at both endpoints of the ability distribution:* $\Delta_t(\underline{\theta}) = \Delta_t(\overline{\theta}) = 0$ *for all* $t \geq 0$.

5 Implementation in a Market Economy

Consider now a market economy with credit markets and taxes. The agent's problem in a market economy with credit markets is to maximize utility subject to a present value budget constraint, taking a tax system as given. We propose a tax system where income taxes depend *only* on current income and schooling subsidies depend *only* on current schooling, i.e. income that could have been earned if the person devoted all her schooling time to work.[12] We will later discuss if this tax system is sufficient to implement the efficient allocations and numerically verify the conditions.

The proposed tax system involves a sequence of income taxes and schooling subsidies $\mathcal{T} = \{T_t\}_{t=0}^{\infty}$ where $T_t : R_+ \times [0,1] \to R$ is a function of only current income and current schooling. The agent's problem in a competitive equilibrium is

$$\max_{\{c_t, l_t, h_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t [c_t - v(l_t + S(h_t, h_{t+1})]$$

subject to the present value budget constraint

$$\sum_{t=0}^{\infty} \beta^t c_t = \sum_{t=0}^{\infty} \beta^t [\theta h_t l_t - T_t(\theta h_t l_t, S(h_t, h_{t+1}))].$$

[12] There may be a discretion in choosing the way schooling subsidies are administered. The subsidies could alternatively depend on the current schooling only, current human capital only or future human capital only. The reason why we prefer the tax system chosen is that it has similar domain to income taxes. Marginal schooling subsidies are then directly comparable to marginal income taxes.

Let λ be the Lagrange multiplier on the budget constraint. Eliminating consumption and writing the problem recursively, one rewrites the agent's problem as follows:

$$\Omega^{\mathcal{T}}(h_0, \theta) = \max_{\{l_t, h_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t [\theta h_t l_t - T_t(\theta h_t l_t, S(h_t, h_{t+1})) - v(l_t + S(h_t, h_{t+1}))]. \tag{15}$$

Let $\psi^{\mathcal{T}} = \{y_t^{\mathcal{T}}, h_{t+1}^{\mathcal{T}}\}_{t=0}^{\infty}$ be the optimal policy functions, where $y_t^{\mathcal{T}} : \Theta \to \mathbf{R}_+$ is the income chosen in period t and $h_{t+1}^{\mathcal{T}} : \Theta \to \mathbf{R}_+$ is the next period human capital chosen in period t. A (partial) equilibrium in the market economy consists of the tax policy $\mathcal{T}$ and the optimal policy functions $\psi^{\mathcal{T}}$ such that the optimal policy functions solve the above problem taking $\mathcal{T}$ as given and the government's present value budget constraint is satisfied.

We will now informally analyze the equilibrium in the market economy. Assume that the tax system is differentiable in both arguments and that both schooling and labor supply are always strictly positive. First-order conditions in labor supply and next period human capital are

$$v_t' = \theta h_t (1 - \frac{\partial T_t}{\partial y}) \tag{16}$$

$$\frac{v_t'}{G_{s_t}} = \beta [v_{t+1}' \frac{G_{h_{t+1}}}{G_{s_{t+1}}} + \theta l_{t+1} (1 - \frac{\partial T_t}{\partial y})] - (\frac{1}{G_{s_t}} \frac{\partial T_t}{\partial s} - \beta \frac{G_{h_{t+1}}}{G_{s_{t+1}}} \frac{\partial T_{t+1}}{\partial s}). \tag{17}$$

It follows from (17) and (13) that the relationship between the marginal schooling subsidy and the human capital wedge is given by

$$\Delta_t(\theta) = -\frac{1}{G_{s_t}} \frac{\partial T_t}{\partial s} + \beta \frac{G_{h_{t+1}}}{G_{s_{t+1}}} \frac{\partial T_{t+1}}{\partial s}.$$

Next proposition states sufficient conditions for the marginal schooling subsidies to be positive.

Proposition 7 *If $\Delta_t(\theta) > 0$ for all $t \geq \bar{t}$ then*

$$\frac{\partial T_{t+1}}{\partial s}(y_t(\theta), s_t(\theta)) < 0 \qquad \textit{for all} \quad t \geq \bar{t}.$$

The Proposition 7 shows that schooling subsidy is positive as long as the human capital wedge continues to be positive for all future periods. Corollary 5 then implies that the marginal schooling subsidy will always be positive in, or nearby, the steady state.

Lemma (6) and Proposition (7) imply immediately that the marginal schooling subsidies must be zero at both ends of the ability distribution:

Lemma 8 *The marginal schooling subsidies are zero at both endpoints of the ability distribution: $\frac{\partial T_{t+1}}{\partial s}(y_t(\underline{\theta}), s_t(\underline{\theta})) = \frac{\partial T_{t+1}}{\partial s}(y_t(\bar{\theta}), s_t(\bar{\theta})) = 0$ for all $t \geq 0$.*

Equations (16) and (17) can define the optimal marginal income taxes and optimal marginal schooling subsidies as a "residual" of the first-order conditions to the individual's distorted problem, whenever the optimal tax system is differentiable. This approach is intuitive, but suffers from the obvious deficiency that the first order conditions (16) and (17) may not be sufficient to implement the optimum. The reason is that the agent may find it optimal to choose income $y_t(\hat{\theta}_1)$ and schooling $s_t(\hat{\theta}_1)$ in period t and income $y_{t+1}(\hat{\theta}_2)$ and schooling $s_{t+1}(\hat{\theta}_2)$ in period $t+1$ for $\hat{\theta}_1 \neq \hat{\theta}_2 \neq \theta$. That is, the agent may behave like a $\hat{\theta}_1$ agent in one period and to behave like a $\hat{\theta}_2$ agent in another period.

We do need to stress that such deviation is only potential. While we cannot rule it out theoretically, it may not appear in the optimum. At the same time, we see a great benefit from using a tax system that is simple enough to be applicable in the real world. Our strategy is therefore to check numerically whether any of these violations happens in the optimum. We will now devise an implementation verification procedure that will verify if the proposed tax system indeed implements the optimum. Only if the implementation verification procedure fails, one will need to devise a more complicated tax system.

5.1 Implementation Verification Procedure

Let ψ^* be the efficient allocation that solves the social planner's problem and $\mathcal{T}^*$ be the tax system defined as follows: If there is an agent such thay $(y, s) = (y_t^*(\theta), s_t^*(\theta))$, define the tax payments $T_t(y, s)$ by integrating the marginal tax rates from (16) into (17).[13].For any other pairs (y, s), let $T_t(y, s) = b$ where b is large enough so that all the agents are discouraged from making such choices. The implementation verification procedure checks if the efficient allocation ψ^* constitutes the market equilibrium under $\mathcal{T}^*$, that is, if $\psi^* = \psi_{\mathcal{T}^*}$.

To numerical determine whether this holds, we formulate the agent's problem in a market economy (15) recursively. Let $\Omega_t^{\mathcal{T}}(h, \theta)$ be the $\theta-$type agent's value of having human capital h at the beginning of period t. The principle of optimality implies that the value function $\Omega_t^{\mathcal{T}} : [0, \bar{h}] \times \Theta \rightarrow \mathbf{R}$ satisfies

$$\Omega_t^{\mathcal{T}^*}(h, \theta) = \max_{y,s}\{y - T_t(y, s) - v(\frac{y}{\theta h} + s) + \beta\Omega_{t+1}^{\mathcal{T}^*}(G(h, s), \theta)\},$$

[13]The intercept is irrelevant since there are no income effects.

where the agent's choice has been reformulated in terms of the income $y = \theta hl$. Let $\{y_t^{\mathcal{T}^*}(h,\theta), s_t^{\mathcal{T}^*}(h,\theta)\}_{t=0}^{\infty}$ be the optimal policy functions.[14] The implementation verification procedure will be successful if for all $t \geq 0$ and for all $\theta \in \Theta$, $|y_t^{\mathcal{T}^*}(h_t^*(\theta),\theta) - y_t^*(\theta)| < \varepsilon$ and $|s_t^{\mathcal{T}^*}(h_t^*(\theta),\theta) - s_t^*(\theta)| < \varepsilon$ for some error tolerance $\varepsilon > 0$.

6 Numerical simulations

In this section we will calibrate the model to the U.S. data. After that, we will compute the optimal policy reform and compare it to the benchmark economy. We will also compare the optimal policy reform to two partial policy reforms, where we only allow the government to change income taxes or schooling.

6.1 Benchmark Economy

We assume that the agents have a CRRA utility $U(u) = \frac{u^{1-\sigma}}{1-\sigma}$, and that $v(n) = \eta\frac{n^{1+\rho}}{1+\rho}$ so that labor supply exhibits a constant Frisch elasticity of $\frac{1}{\rho}$ (in a static model with no human capital). We set $\sigma = 1$ and $\eta = 1$ and $\rho = 4$, so that the elasticity of labor supply is 0.25. The time period is one year and so the discount factor β is set equal to 0.96. We also assume that the human capital production function is Cobb-Douglas, $G(h,s) = (1-\delta)h + \delta h^{\alpha}s^{1-\alpha}$. We assume the Ben-Porath specification with $\alpha = 0.5$. The depreciation δ is set to 0.04, which is roughly in the middle of the available evidence.[15]

The empirical distribution of earnings is obtained from the 1992 tax return data. We

[14] Since one can costlessly penalize the agent for choosing income outside the range of y_t^* and the forgone income outside the range of z_t^* where z_t^* is defined by $z_t^*(\theta) = \theta h_t^*(\theta)S(h_t^*(\theta), h_{t+1}^*(\theta))$, it is innocuous restrict the agent's choice of y and z to the range of y_t^* and z_t^*. This feature is very convenient numerically.

[15] See Browning, Hansen and Heckman [5] or Trostel [30].

calibrate the distribution of abilities in such a way that the steady state distribution of earnings in the model matches the empirical distribution of earnings in a model where the agents face a flat income tax of 40% and no schooling subsidies. By assuming that there are no schooling subsidies in the U.S., we provide an upper bound on the possible welfare gains from optimal schooling policies. Finally, the government expenditures are assumed to be equal to 25% of the aggregate income.

Since the data and hence the calibrated abilities levels not very smooth, we use a double Pareto-Lognormal distribution to approximate the empirical distribution of abilities and discretize it by a vector of $I = 300$ gridpoints.[16]

6.2 The Optimal Policy Reform

In this subsection, we analyze the efficient allocations and optimal policies in the calibrated economy. When solving for the efficient allocations we follow the following numerical procedure. We fix the utility allocation u and the Lagrange multiplier on the resource constraint λ. We then compute the cumulative distortion function $X_{u,\lambda}$ and adjust u_0 until this function satisfies $X_{u,\lambda}(\overline{\theta}) = 0$. We solve the dynamic program (11) for each agent and generate the sequence of optimal human capital and labor supply allocations for $\bar{t} = 250$ periods.[17] The envelope condition (4) is then used to compute a new utility allocation Tu. We repeat the procedure until $\|Tu - u\| \leq \varepsilon$ for some error tolerance $\varepsilon > 0$. After that,we check if the resource constraint holds with equality (up to an error tolerance). If not, we update the Lagrange multiplier λ until the resource constraint is satisfied with equality.

[16] See Jorgensen and Reed [17] for details of the Pareto-Lognormal distribution.
[17] The economy is approximately in a steady state in period T.

Figure 1 plots the optimal marginal tax rates in periods 1, 50, and 250. It plots them against the ability level of the agents.[18] It shows that marginal income tax rates change extremely little over time, despite the fact that, as we will see later, there are significant changes in human capital across population. The marginal income taxes are zero at both endpoints of the distribution, as predicted by the theory. The shape of the marginal income tax schedule is primarily determined by the shape of the ability distribution.

INSERT FIGURE 1

The intertemporal human capital wedge is shown in Figure 2. The wedge is zero at both endpoints and strictly positive between them for all time periods.[19] It is smaller than the marginal income taxes. While the marginal income taxes decrease for most of the population, the intertemporal wedge tends to increase. Both of these patterns follow from the fact that it is in general more costly for the social planner to distort the behavior of higher ability agents, rather than to distort the behavior of lower ability agents. The marginal income taxes are thus higher for low ability people because that allows the government to extract resources from high ability people without distorting their incentives. Similarly, the distortions of human capital investment become relatively more important for high ability agents. Hence their human capital wedge, which encourages them to invest in human capital, becomes larger.

INSERT FIGURE 2

[18]To get some idea about the ability distribution, note that the expected value of theta is 0.66 and 99th percentile is represented by $\theta = 2.25$.

[19]One can verify that the conditions of Proposition (4) are satisfied in all time periods and so the intertemporal wedge must be positive.

Figure 3 shows the marginal schooling subsidies. Compared to the marginal income taxes, the marginal schooling subsidies are quite small. Similarly to the marginal income taxes they tend to change very little over time and for most of the population they lie between 0 and 7%. For most of the population they are increasing, although they decrease at the top.

INSERT FIGURE 3

Figure 4 shows the aggregate dynamics of the economy. Aggregate output initially decreases as people spend their time by schooling rather than working. It exceeds the initial level after about 20 periods and is 6.1% higher in the new steady state. Interestingly, the aggregate consumption is lower than the pre-reform aggregate consumption in all periods. This, together with fairly low aggregate gain in output indicates that most of the welfare gains comes from a more efficient reallocation of labor supply and schooling across people that cancels out in the aggregate, rather than from an increase of labor supply and schooling.

INSERT FIGURE 4

6.2.1 Implementation Verification Procedure Results

We compute the tax policy $\{T_t\}_{t=0}^{\infty}$ by numerically integrating the marginal income taxes $\frac{\partial T_t}{\partial y}$ and $\frac{\partial T_t}{\partial s}$ obtained from equations (16) and (17). Because the distribution of the ability parameter θ has been discretized by a vector of $I = 300$ points, the implementation verification procedure will be successful if the optimal policy functions $y_t^{\mathcal{T}^*}(h_t^*(\theta_i), \theta_i)$ and $s_t^{\mathcal{T}^*}(h_t^*(\theta_i), \theta_i)$ are exactly equal to their "target" values $y_t^*(\theta_i)$ and $s_t^*(\theta_i)$ for all $t = 1..T$ and $i = 1..I$.

We compute the implementation verification procedure for all $\bar{t}$ periods and for every every agent We find that for all the periods and all the agents the efficient allocation passes the implementation verification procedure.

6.3 Other Policy Reforms

In this subsection we analyze alternative, more restricted, policy reforms and compare them with the optimal policy reform analyzed in previous section. We study two alternatives: a reform where the government continues not using schooling subsidies at all, but chooses income taxes optimally and a reform where the government chooses schooling subsidies optimally, but continues using a flat income tax of 40%.

Figure 5 compares optimal marginal income taxes in the optimal policy reform (when schooling subsidies are set optimally) and in the first restricted policy reform (when schooling subsidies are zero). Similarly, Figure 6 compares optimal marginal schooling subsidies in the optimal policy reform and in the second restricted policy reform (when income taxes are flat). To compare the dispersion of marginal rates over time, we plot them for all 250 time periods.

INSERT FIGURE 5

Two insights can be obtained from Figure 5 and 6. First, marginal income taxes are significantly more volatile when no schooling subsidies are used. In particular, they tend to decrease over time. Only when schooling subsidies and income taxes are both set optimally we see very small variations in income taxes over time. Second, as shown in Figure 6, when income tax is flat, marginal schooling subsidies are higher for people with very high abilities, but lower for people with very low abilities. This is caused by the fact that for high ability people the flat income tax rate is too high compared to the optimal marginal income tax rate. To partially mitigate this effect, the government encourages human capital accumulation at a greater rate. The opposite logic applies for the lower half of the population where the optimal marginal income tax rate is higher than the flat income tax rate. In fact, when the income tax is flat, the schooling subsidies turn out to be negative for agents with very low abilities. The government thus tries to extract additional resources from the low ability agents. Note also that when the marginal income tax is flat, the optimal schooling subsidies are nonzero even at the upper endpoint of the distribution. This is so because the flat tax rate distorts incentives even at the upper endpoint of the distribution and so positive schooling subsidy is called for to partially "offset" the inefficiency.

INSERT FIGURE 6

Table 1 compares the aggregate effects of all 3 reforms with the pre-reform values. The first four columns show how steady state values of aggregate output, consumption, labor supply and schooling[20] change from their calibrated values. The aggregate output gain from the optimal tax reform is slightly larger than the output gain from the first restricted reform when no schooling subsidies are used, but is significantly smaller than the output gain from the second partial reform when only schooling subsidies are used. Surprisingly, the aggregate labor supply is the smallest in the optimal tax reform when it decreases by 5.3%. In both

[20]The Cobb-Douglas production function implies that steady state changes in human capital are identical to steady state changes in schooling.

partial reforms the aggregate labor supply changes very little. Aggregate schooling changes significantly whenever schooling subsidies are used and stays almost the same in the second partial reform when only income tax changes. Schooling subsidies are thus essential to increase aggregate schooling.

The fifth column shows the relative welfare gains from the tax reforms, measured in terms of consumption equivalents.[21] The welfare gain from the introduction of both schooling subsidies and optimal income taxes (10%) is almost the same as the welfare gain from the introduction of optimal income taxes only (9.7%). Thus, as long as the income taxes are set optimally, the additional welfare gain from schooling subsidies is very small. If, however, the income tax function is restricted to be flat then the welfare gains from schooling subsidies are large (8.8%). When one compares the two restricted reforms directly, income tax appears to be a more powerful instrument to increase welfare, but the difference is small.

INSERT TABLE 1

Figure 7 investigates further the differences between all three reforms by comparing the cumulative welfare gains. It plots the average welfare gains of a population below a given ability level.[22] The optimal reform and the first restricted reform are almost equivalent. These reforms are particularly attractive to the low ability people who experience large welfare gains. The second restricted reform when only schooling subsidies change is different in that welfare gains are more evenly distributed across population. The welfare gains of

[21] If u^0 is the calibrated utility distribution and u^* is the optimal distribution, then the relative welfare gain γ is a solution to $\int_\Theta U[u^0(1+\gamma)]fd\theta = \int_\Theta U[u^*]fd\theta$.

[22] If u^0 is the calibrated utility distribution and u^* is the optimal distribution, then the cumulative relative welfare gain is given by $\int_{\underline{\theta}}^{\theta}[U(u^0) - U(u^*)]fd\tilde{\theta}/(1-F(\theta))$.

low ability people under the first two reforms are large because under the optimal income tax function the government gives all agents large lump sum transfers. This is possible since at the same time the government imposes high marginal tax rates on low ability people, which extracts resources from high ability people without distorting their incentives. The flat tax is not as efficient in extracting resources. Although schooling subsidies to some extent substitute for the income tax, the substitution is not perfect.

INSERT FIGURE 7

7 Conclusions

We have set up a dynamic private information model with endogenous accumulation of observable human capital and analyzed the optimal income taxes and schooling policies. We found that under certain plausible conditions, and certainly in a steady state, the optimal schooling subsidies and the intertemporal human capital wedge are both positive. We have calibrated the model according to the U.S. economy and found that the marginal schooling subsidies are between 0-7% for almost all of the population for almost all time periods. Overall the magnitude of marginal schooling subsidies is significantly smaller than the magnitude of marginal income taxes.

We have also analyzed various partial reforms where either the income taxes or schooling subsidies are exogenously given. We find that, as long as the income taxes are set optimally, the welfare gain from the introduction of schooling subsidies is fairly small. On the other hand, schooling subsidies become much more significant if income taxes are not set optimally.

References

[1] ABRAHAM, A. and N. PAVONI, (2004). “Efficient Allocations with Moral Hazard and Hidden Borrowing and Lending.” University of Rochester, unpublished manuscript.

[2] ALBANESI, S. and Ch. SLEET (2003). “Dynamic optimal Taxation with Private Information.” Forthcoming, *The Review of Economic Studies*.

[3] ANDOLFATTO, D., Ch. FERRALL and P. GOMME (2000). ”Life Cycle Learning, Earning, Income and Wealth”. Simon Fraser University, unpublished manuscript.

[4] BOVENBERG, A. L. and B. JACOBS (2003). ”Redistribution and Education Subsidies are Siamese Twins” (University of Amsterdam/Tilburg, mimeo)

[5] BROWNING, M., HANSEN, L.P. and HECKMAN, J.J. (2000). “Micro Data and General Equilibrium Models.” (In J. B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics* Vol 1a, Chapter 8).

[6] CARNEIRO, P. and J. J. HECKMAN (2004). “Human capital policy.”In J. Heckman, A. Krueger, *Inequality in America: What Role for Human Capital Policies*, MIT Press.

[7] DIAMOND, P. (1998). “Optimal Income Taxation: an Example with a U-Shaped Pattern of Optimal Marginal Tax Rates.” *The American Economic Review,* **88**, 83-95.

[8] DIAMOND, P. and MIRRLEES, J.A. (2002). “Optimal Taxation and the Le Chatelier Principle.” MIT, unpublished manuscript.

[9] FARHI, E. and WERNING, I. (2005). “Inequality, Social Discounting and Progressive Estate Taxation. MIT, unpublished manuscript.

[10] GOLOSOV, M., N. KOCHERLAKOTA and TSYVINSKI, A. (2003). "Optimal Indirect and Capital Taxation", *Review of Economic Studies*, **70**, 569-588.

[11] GOLOSOV, M. and TSYVINSKI, A. (2006). "Optimal Taxation with Endogenous Insurance Markets." NBER Working Paper #11185.

[12] GROCHULSKI, B. and T. PISKORSKI (2005). 'Optimal Wealth Taxes with Risky Human Capital', NYU, unpublished manuscript.

[13] HEATHCOTE, J., K. STORESLETTEN and G. L. VIOLANTE (2004). "The Macroeconomic Implications of Rising Wage Inequality in the United States.", Georgetown University, unpublished manuscript.

[14] HECKMAN, J. J. (1975). "Estimates of a Human Capital Production Function Embedded in a Life-Cycle Model of Labor Supply." (In N. Terlekyj, ed., *NBER Studies in Income and Wealth*, **40**, New York, Columbia University Press).

[15] HECKMAN, J. J., L. LOCHNER and Ch. TABER (1998). "Tax Policy and Human-Capital Formation." *The American Economic Review*, **88**, Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association, 293-297.

[16] HUGGETT, M., G. VENTURA and A. YARON (2003). "Human Capital and Earnings Distribution Dynamics" (Mimeo, Georgetown University).

[17] JORGENSEN, M. and REED, W.J. (2004). "The Double Pareto-Lognormal Distribution - A New Parametric Model for Size Distributions." *Com. Stats -Theory & Methods*, **33**.

[18] KAPICKA, M. (2006). ”Optimal Taxation with Endogenous Human Capital Accumulation and Limited RecordKeeping”, *Review of Economic Dynamics*, forthcoming.

[19] KAPICKA, M. (2006). ”The Dynamics of Optimal Taxation when Human Capital is Endogenous.” Unpublished Manuscript, UC Santa Barbara.

[20] KOTLIKOFF, L., SMETTERS, K. and WALLISER, J. (2001). “Finding a Way Out of America’s Demographic Dilemma.” Boston University, unpublished manuscript.

[21] KOCHERLAKOTA, N. (2005). “Zero Expected Wealth Taxes: A Mirrlees Approach to Dynamic Optimal Taxation.” *Econometrica*, **73**, 1587-1621.

[22] KOCHERLAKOTA, N. (2004). “Figuring out the impact of hidden savings on optimal unemployment insurance.” *Review of Economic Dynamics*, **7**, 541-554.

[23] MIRRLEES, J.A. (1971). “An Exploration in the Theory of Optimum Income Taxation.” *Review of Economic Studies,* **38**, 175-208.

[24] MIRRLEES, J.A. (1976). “Optimal Tax Theory: A Synthesis.” *Journal of Public Economics*, **6,** 327-358.

[25] MIRRLEES, J.A. (1986). “The Theory of Optimal Taxation.” (In K.J. Arrow and M.D. Intriligator, eds., *Handbook of Mathematical Economics*, vol. III, chapter 24).

[26] SAEZ, E. (2001). “Using Elasticities to Derive Optimal Income Tax Rates.” *Review of Economic Studies,* **68**, 205-229.

[27] STOKEY, N.l., LUCAS, R.E. and PRESCOTT R.C. Recursive Methods in Economic Dynamics. Cambridge, Harvard University Press.

[28] TABER, Ch. (2001). "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies* **68**, 665-691.

[29] TABER, Ch. (2002). "Tax Reform and Human Capital Accumulation: Evidence from an Empirical General Equilibrium Model of Skill Formation." *Advances in Economic Analysis and Policy* **2**, No. 1, Article 3.

[30] TROSTEL, P A. (1993). "The Effect of Taxation on Human Capital." *Journal of Political Economy*, **101**, 327-350.

[31] WERNING, I. (2001). "Dynamic Optimal Taxation." MIT, unpublished manuscript.

8 Appendix

Proof of Lemma 1. The proof is similar to the proof of [25] that, in a static environment, increasing income is sufficient for the first order approach to be valid. Suppose that an allocation satisfies (4). Let $\hat{\theta} < \theta$. (4) implies that

$$
\begin{aligned}
V_\psi(h_0, \theta; \theta) - V_\psi(h_0, \hat{\theta}; \hat{\theta}) &= \int_{\hat{\theta}}^{\theta} \sum_{t=0}^{\infty} \beta^t v'[\frac{y_t(\varepsilon)}{\varepsilon h_t(\varepsilon)} + S(h_t(\varepsilon), h_{t+1}(\varepsilon)] \frac{y_t(\varepsilon)}{\varepsilon h_t(\varepsilon)} \frac{d\varepsilon}{\varepsilon} \\
&\geq \int_{\hat{\theta}}^{\theta} \sum_{t=0}^{\infty} \beta^t v'[\frac{y_t(\hat{\theta})}{\varepsilon h_t(\hat{\theta})} + S(h_t(\hat{\theta}), h_{t+1}(\hat{\theta})] \frac{y_t(\hat{\theta})}{\varepsilon h_t(\hat{\theta})} \frac{d\varepsilon}{\varepsilon} \\
&= V_\psi(h_0, \theta; \hat{\theta}) - V_\psi(h_0, \hat{\theta}; \hat{\theta}).
\end{aligned}
$$

The first equality follows from the fundamental theorem of calculus. The inequality follows from monotonicity of $\frac{y_t(\theta)}{h_t(\theta)}$ and $S(h_t(\theta), h_{t+1}(\theta)$ and the fact that $\varepsilon \geq \hat{\theta}$. Finally, the last equality applies (4). The proof is similar for $\hat{\theta}. > \theta$. ■

Proof of Lemma 2. $X_{u,\lambda}(\overline{\theta}) = 0$ is obvious. $X_{u,\lambda}(\underline{\theta}) = 0$ follows from (8). As follows from (4), u must be increasing in θ. Since U is strictly concave, U' is decreasing in θ. Equation (8) then implies that $U'(u)-\lambda$ is first positive and then negative. Consequently, $\int_{\underline{\theta}}^{\theta}[U'(u)-\lambda]fd\varepsilon$ is always positive. ■

Proof of Theorem 3. Necessity of the conditions is obvious. For sufficiency, let $\psi^* = \{u^*(\theta), l_t^*(\theta), h_{t+1}^*(\theta)\}$, together with $\hat{\lambda}^*$ and $\mu^* = \lambda^* - U'(u^*)$ be an allocation that satisfies the conditions of the theorem. Let $\psi = \{u(\theta), l_t(\theta), h_{t+1}(\theta)\}$ be an alternative allocation that solves the relaxed social planner's problem. I will prove sufficiency by showing that the allocation ψ^* delivers the value of the social welfare function expected utility at least as high as the one delivered by ψ.

Since U is increasing and concave, We have $U(u^*) - U(u) \geq U'(u^*)(u^* - u)$. Hence

$$\int\limits_{\Theta}[U(u^*) - U(u)]fd\theta \geq \int\limits_{\Theta} U'(u^*)(u^* - u)fd\theta$$

and so it will suffice to show that the right hand side of the inequality is positive. To show this, substract first the two corresponding envelope conditions. We have

$$u^*(\theta) - u(\theta) = u^*(0) - u(0) + (1-\beta)\int_{\underline{\theta}}^{\theta}\sum_{t\geq 0}^{\infty}\beta^t(v_t'^*\frac{l_t^*}{\varepsilon} - v_t'\frac{l_t}{\varepsilon})d\varepsilon.$$

Multiply both sides by $\mu^*(\theta)f(\theta)$ and integrate over Θ :

$$\int\limits_{\Theta}\mu^*(u^* - u)fd\theta = \int\limits_{\Theta}\mu^* f\int_{\underline{\theta}}^{\theta}\sum_{t\geq 0}^{\infty}\beta^t(v_t'^*\frac{l_t^*}{\varepsilon} - v_t'\frac{l_t}{\varepsilon})d\varepsilon$$

since $[u^*(0) - u(0)] \int_\Theta \mu^* f d\theta = 0$. Reversing the order of integration, we get

$$\int_\Theta \mu^*(u^* - u) f d\theta = (1-\beta)\lambda^* \int_\Theta X^*(\theta)\{\sum_{t\geq 0}^{\infty} \beta^t (v_t'^* \frac{l_t^*}{f\theta} - v_t' \frac{l_t}{f\theta})\} f d\theta$$

$$= (1-\beta)\lambda^* \int_\Theta \sum_{t\geq 0}^{\infty} \beta^t \{[\theta h_t l_t + v_t - v_t' l_t \frac{X^*(\theta)}{f\theta}] - [\theta h_t^* l_t^* + v_t^* - v_t'^* l_t^* \frac{X^*(\theta)}{f\theta}]\} f d\theta$$

$$+ (1-\beta)\lambda^* \int_\Theta \sum_{t\geq 0}^{\infty} \beta^t [\theta h_t^* l_t^* + v_t^* - (\theta h_t l_t + v_t)] f d\theta$$

$$\leq \lambda^* \int_\Theta (u^* - u) f d\theta.$$

The inequality follows from two facts. First, $\{l_t^*(\theta), h_{t+1}^*(\theta)\}$ maximizes individual's distorted problem and so the first expression on the right hand side of the second equality is nonpositive. Second, the resource constraint holds with equality and so the second expression on the right hand side of the second equality is equal to $\lambda^* \int_\Theta (u^* - u) f d\theta$. We will now use the fact that $\mu^* = \lambda^* - U'(u^*)$ and so

$$\int_\Theta [\lambda^* - U'(u^*)](u^* - u) f d\theta \leq \lambda^* \int_{\underline{\theta}}^{\theta} (u^* - u) f d\theta.$$

Upon cancelling terms and rearranging the terms, we get that

$$\int_\Theta U'(u^*)(u^* - u) f d\theta \geq 0,$$

which completes the proof. ■

Proof of Proposition 7. In a steady state, the relationship between the wedge and the marginal tax on schooling is thus given by $\Delta = -\frac{1-\beta G_h}{G_s} \frac{\partial T}{\partial s}$. It can be shown that $1 - \beta G_h > 0$

and so $\frac{\partial T}{\partial s} < 0$. The same result applies near the steady state. Suppose now that $\frac{\partial T_{t+1}}{\partial s} < 0$. Using the backward induction, we have

$$\frac{1}{G_{s_t}}\frac{\partial T_t}{\partial s} = -\Delta_t + \beta \frac{G_{h_{t+1}}}{G_{s_{t+1}}}\frac{\partial T_{t+1}}{\partial s} < 0$$

and so $\frac{\partial T_t}{\partial s} > 0$. This result applies as long as $\Delta_t(\theta) > 0$, which is, for all $t \geq \bar{t}$. ■

Figure 1

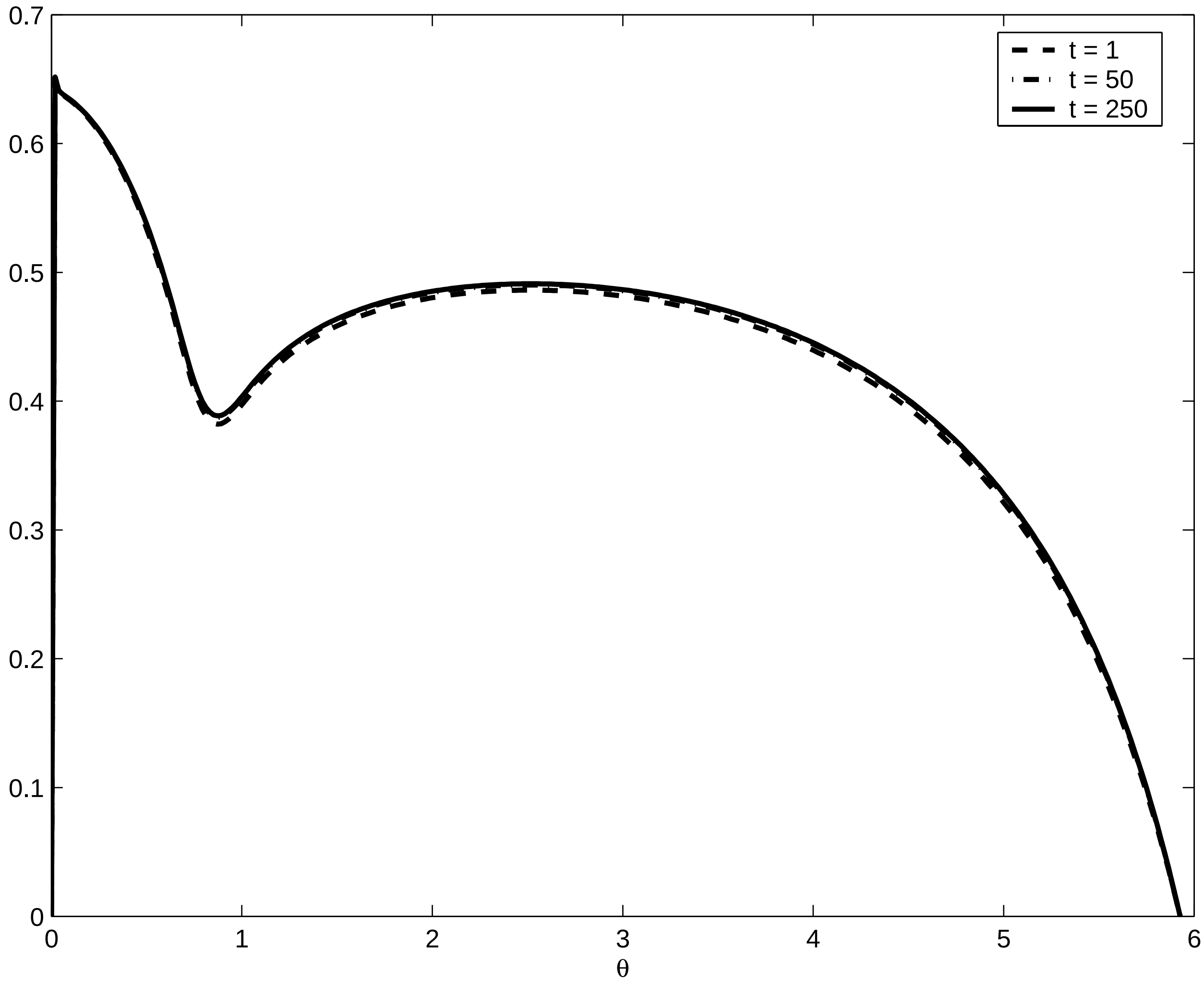


Figure 2

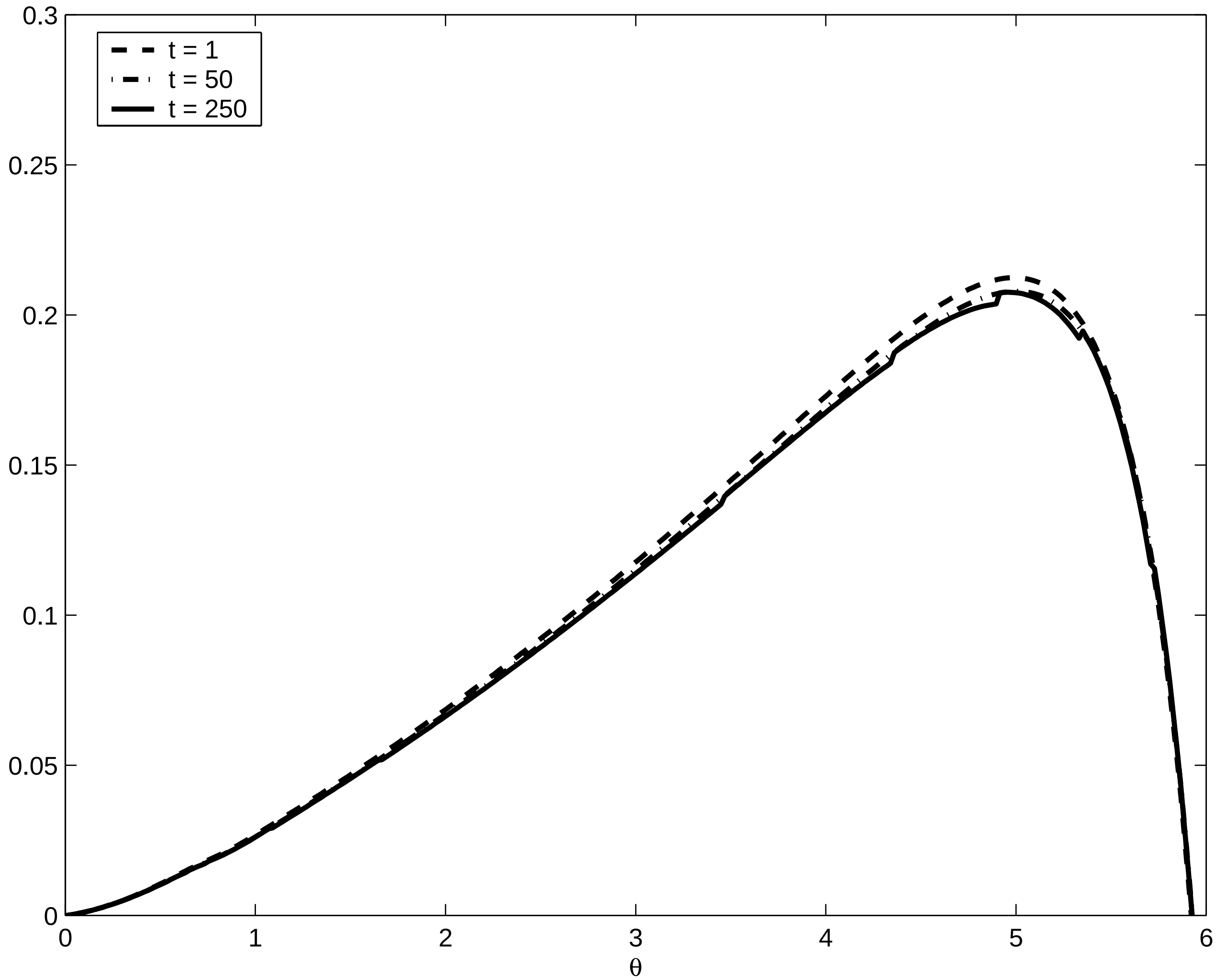


Figure 3

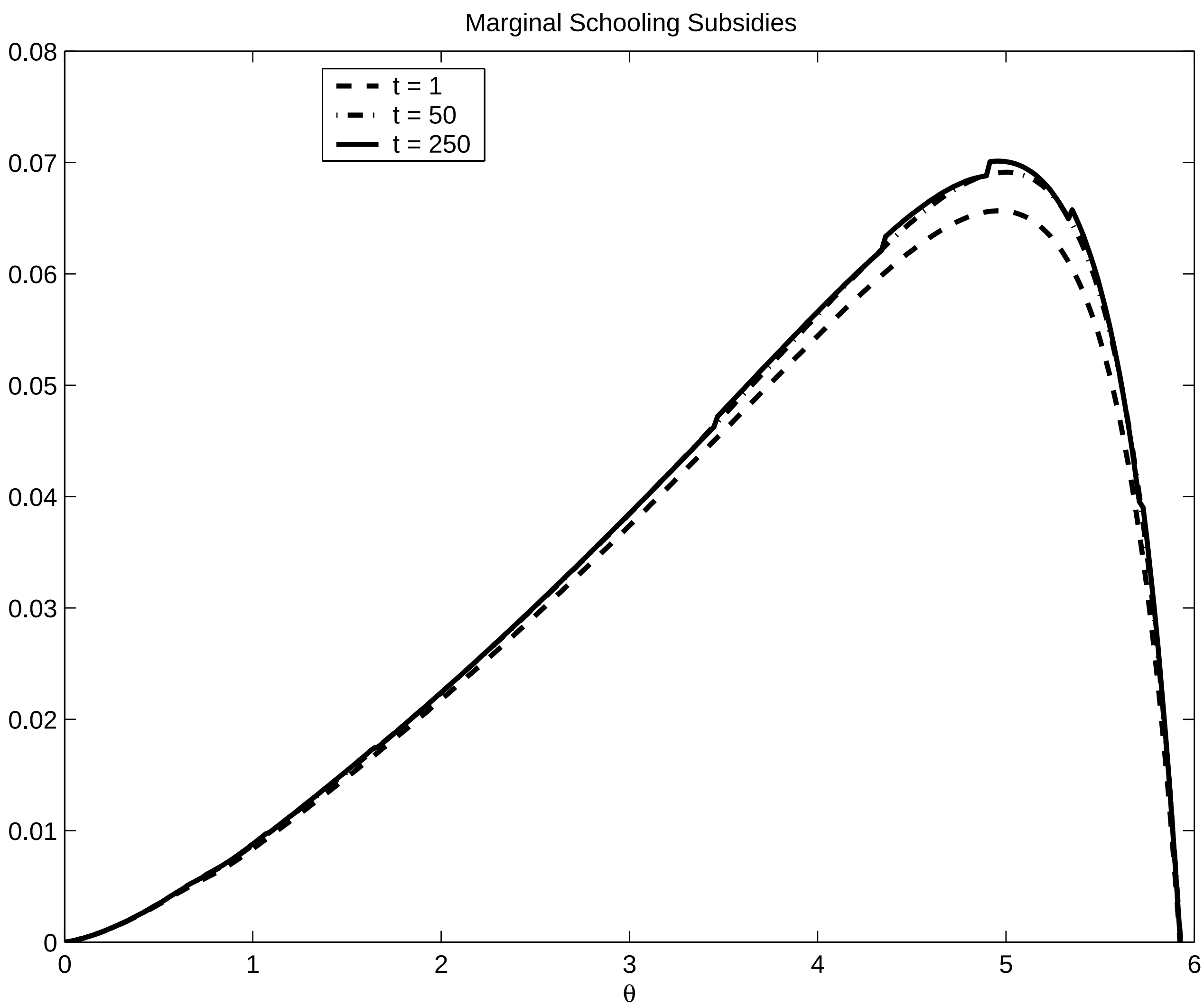

Marginal Schooling Subsidies
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
1
2
3
4
5
6
θ
t = 1
t = 50
t = 250

Figure 4

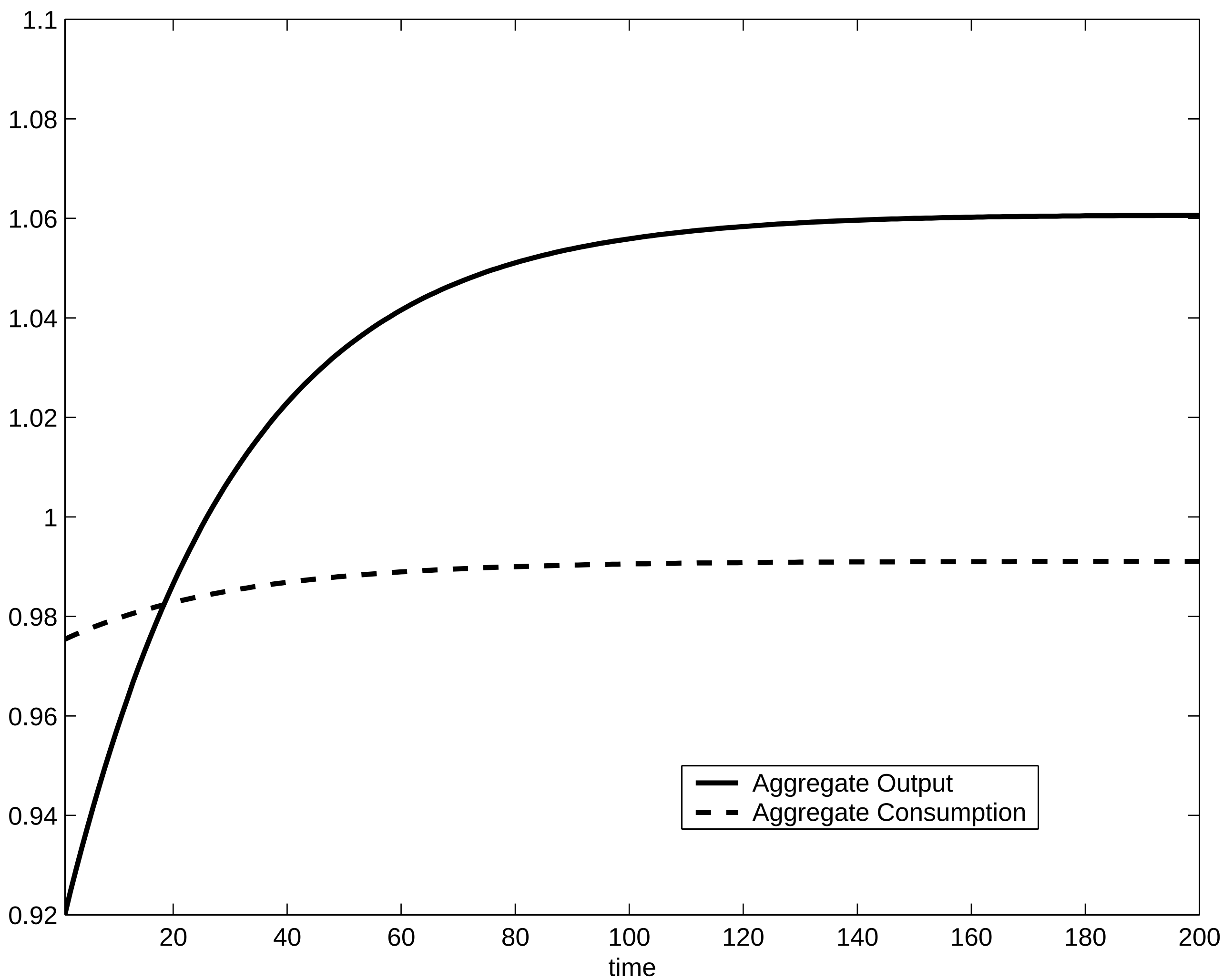
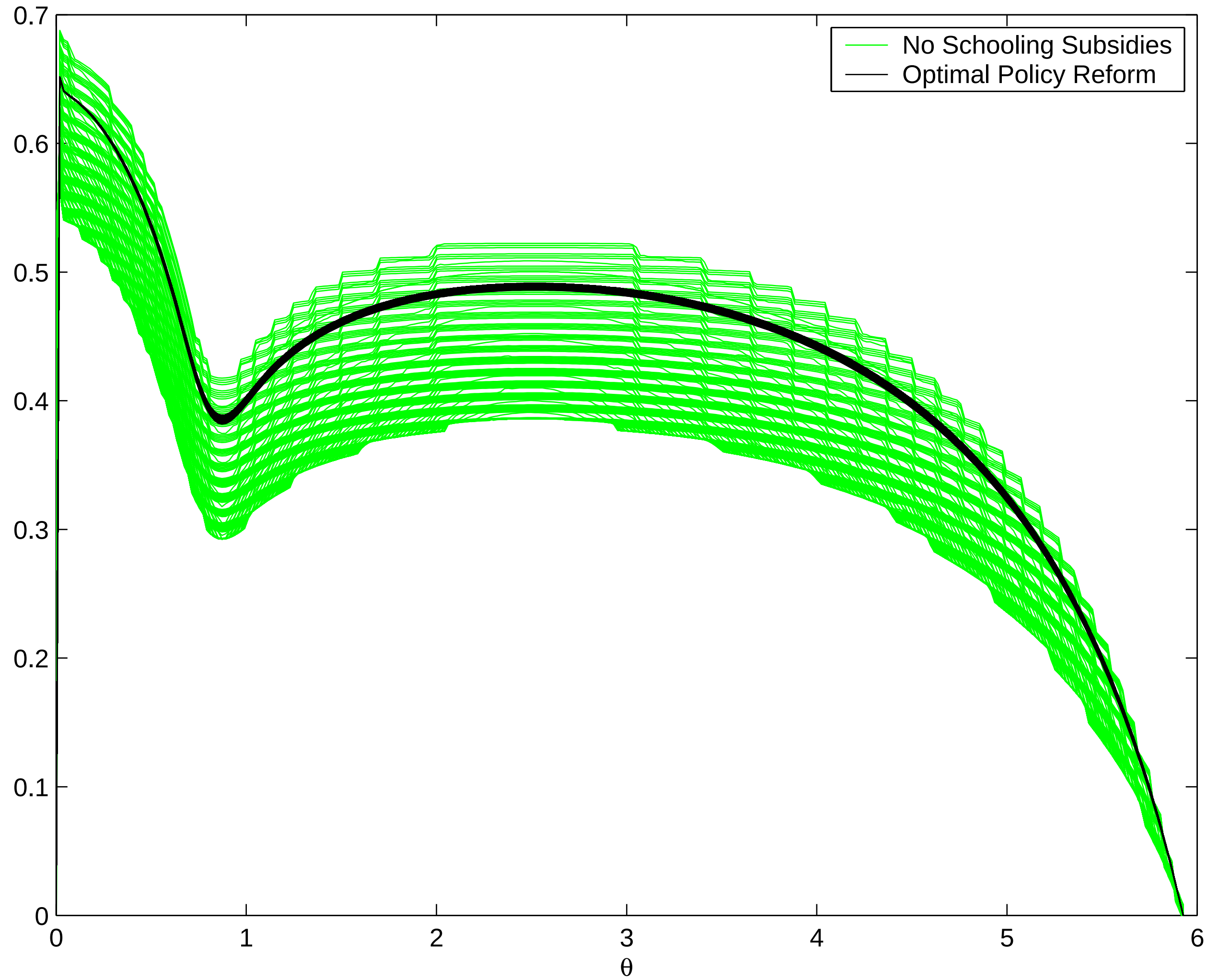
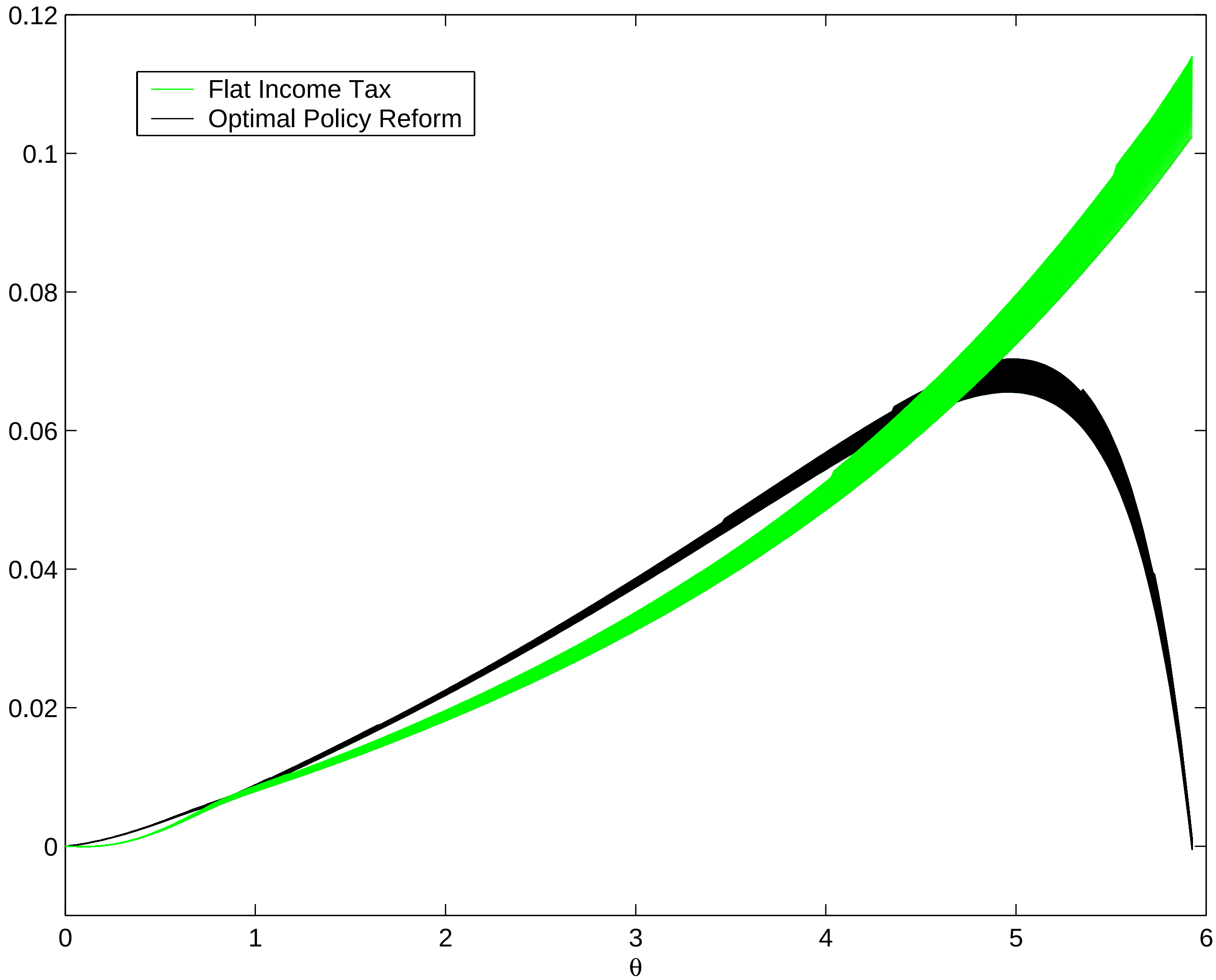

1.1
1.08
1.06
1.04
1.02
1
0.98
0.96
0.94
0.92
20
40
60
80
100
120
140
160
180
200
time
Aggregate Output
Aggregate Consumption

Figure 5



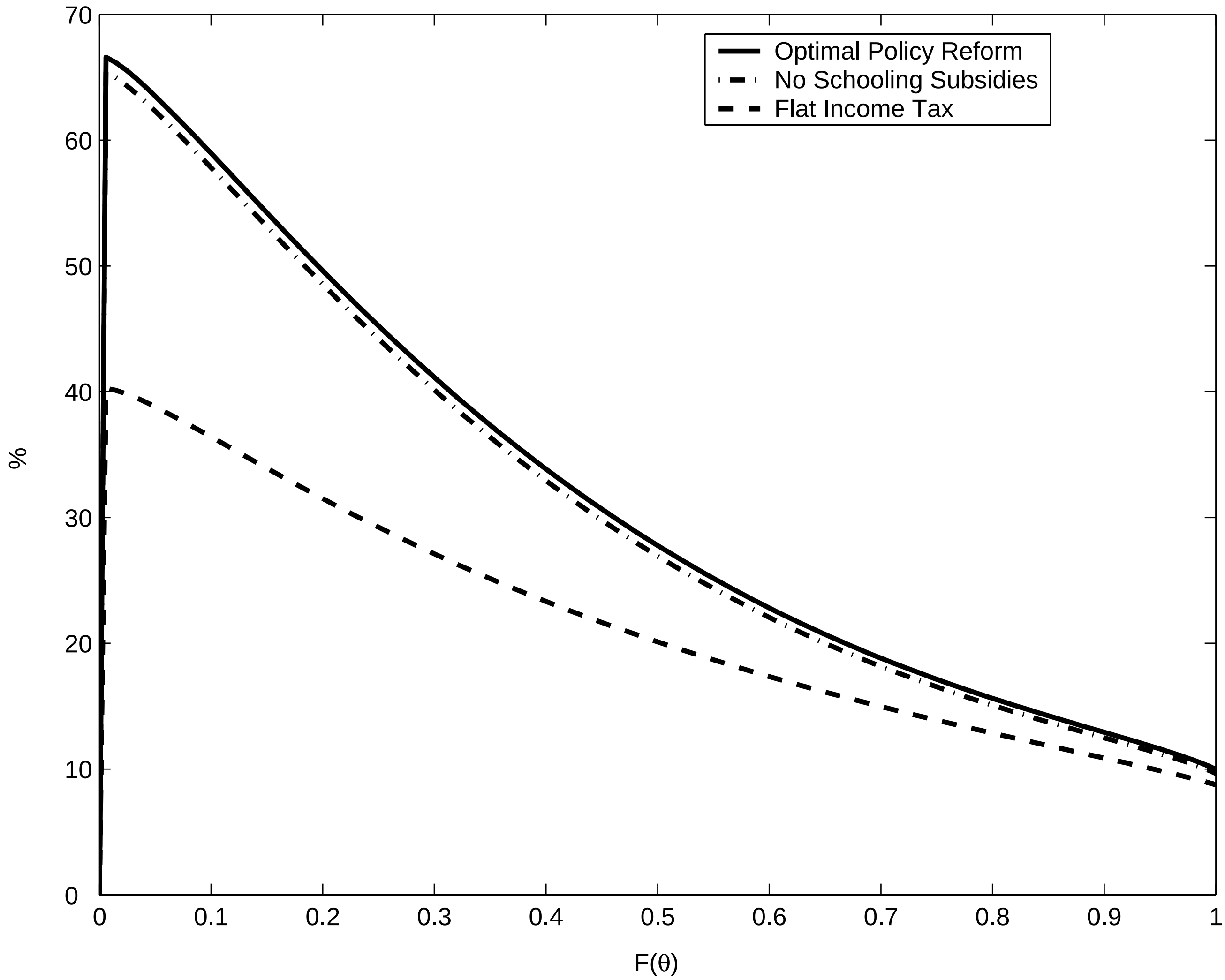
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
θ
No Schooling Subsidies
Optimal Policy Reform

Figure 6



0.12
0.1
0.08
0.06
0.04
0.02
0
Flat Income Tax
Optimal Policy Reform
0
1
2
3
4
5
6
θ

Figure 7


70
60
50
40
30
20
10
0
%
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
F(θ)
Optimal Policy Reform
No Schooling Subsidies
Flat Income Tax

Table

	Output	Cons.	Labor Supply	Schooling	Welfare
Optimal Tax + Optimal Subs.	+6.1%	-1.2%	-5.3%	+8.0%	+10.0%
Optimal Tax + No Subs.	+5.6%	+1.9%	+0.8%	+0.8%	+9.7%
Flat Tax + Optimal Subs.	+9.9%	+3.1%	+0.0%	+8.2%	+8.8%

Table1: Steady State Aggregate Quantities and Welfare, Changes from Pre-Reform Values